

Servicios de Nube - Azure 01

Bases de Datos Avanzadas

Universidad del Rosario

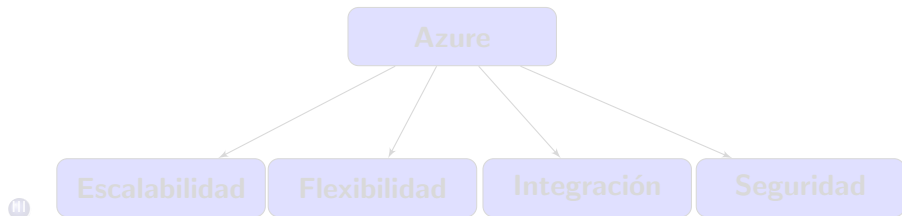
Maestría MACC

- 1 Azure
 - Propiedades de Azure
- 2 Azure Storage Explorer
- 3 Data Lake
 - Tratamiento a Grandes Volúmenes de Datos
 - Diferencias con Data Warehouse
- 4 Azure Data Factory
 - Usos de Data Factory

Características



- 1 Servicio tipo nuble/*cloud*.
- 2 Creación e implementación de aplicaciones y servicios.

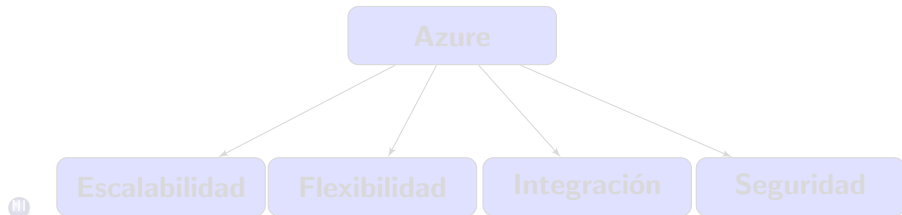


- Integración con Office 365, Power BI y Visual Studio.

Características



- ❶ Servicio tipo nuble/*cloud*.
- ❷ Creación e implementación de aplicaciones y servicios.

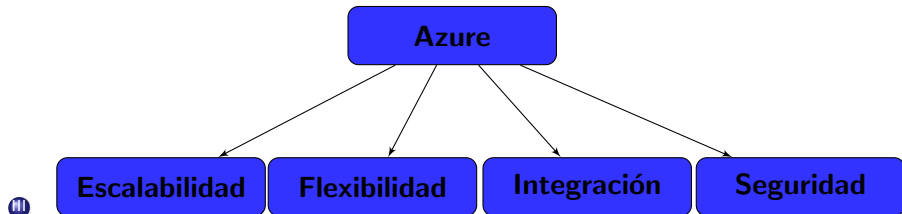


• Integración con Office 365, Power BI y Visual Studio.

Características

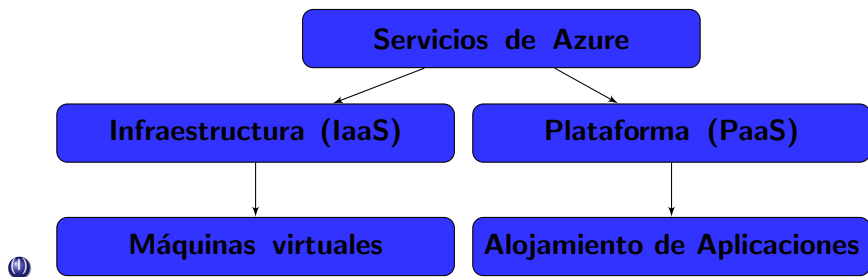


- ❶ Servicio tipo nuble/*cloud*.
- ❷ Creación e implementación de aplicaciones y servicios.



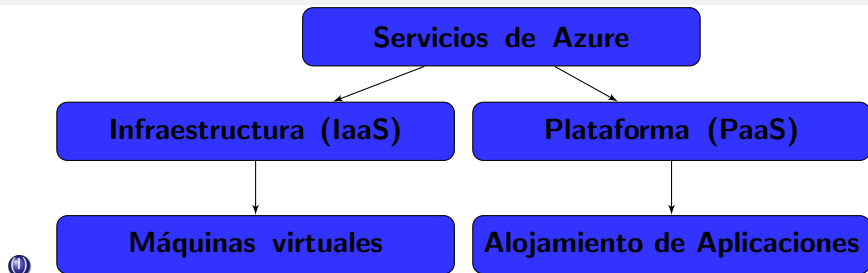
- *Integración con Office 365, Power BI y Visual Studio.*

Escalabilidad

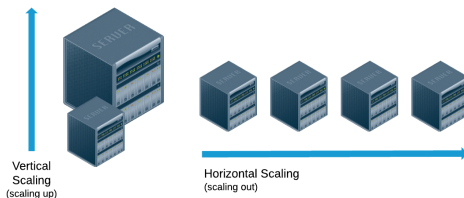


Tipos de escalabilidad: *horizontal y vertical*.

Escalabilidad

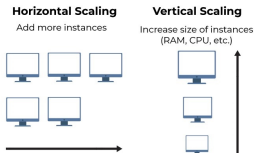


Tipos de escalabilidad: *horizontal* y *vertical*.



Servicios para Escalabilidad Horizontal y Vertical

Horizontal Scaling vs. Vertical Scaling



Horizontal:

- Azure Virtual Machines, Azure Kubernetes Service (AKS), Service Fabric, Azure App Service.
- *Agregar más servidores.*

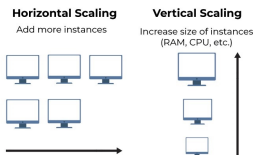


Vertical:

- Azure VM Scale, Azure SQL Database.
- *Potenciar un servidor.*

Servicios para Escalabilidad Horizontal y Vertical

Horizontal Scaling vs. Vertical Scaling



Horizontal:

- Azure Virtual Machines, Azure Kubernetes Service (AKS), Service Fabric, Azure App Service.
- *Agregar más servidores.*



Vertical:

- Azure VM Scale, Azure SQL Database.
- *Potenciar un servidor.*

Flexibilidad y Seguridad

I **Flexibilidad:**

II Servicios de Plataforma como servicio (Paas).

- .NET, Java, Python, Node.js, MySQL, etc.

III Servicios de Infraestructura (IaaS).

- VMs, contenedores y servicios de aplicaciones.

IV **Seguridad:**

- Control de acceso basado en roles (RBAC).
- Encriptación, seguridad de red.

V **Herramientas adicionales:**

- Azure Key Vault, Azure Active Directory, etc.

Flexibilidad y Seguridad

I Flexibilidad:

II Servicios de Plataforma como servicio (Paas).

- .NET, Java, Python, Node.js, MySQL, etc.

III Servicios de Infraestructura (IaaS).

- VMs, contenedores y servicios de aplicaciones.

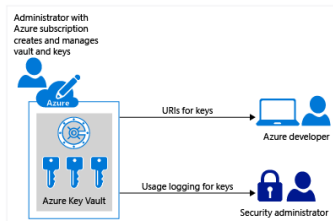
IV Seguridad:

- Control de acceso basado en roles (RBAC).
- Encriptación, seguridad de red.

V Herramientas adicionales:

- **Azure Key Vault**, Azure Active Directory, etc.

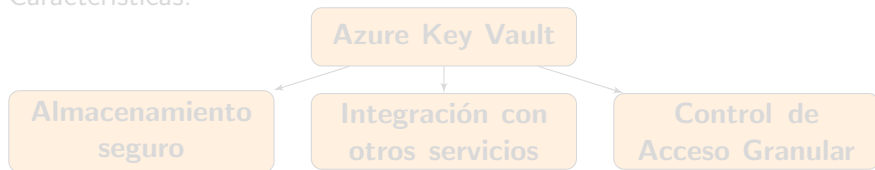
Seguridad: Azure Key Vault



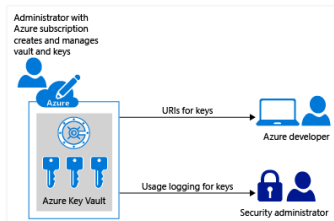
① Administración y creación de claves y secretos.

- *Secreto*: información confidencial usada en una aplicación o servicio (p. ej., certificado de cliente).

② Características:

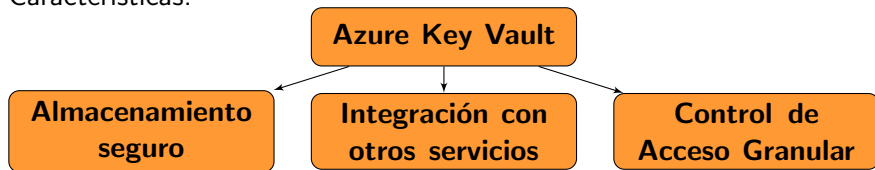


Seguridad: Azure Key Vault



- ① Administración y creación de claves y secretos.
 - *Secreto*: información confidencial usada en una aplicación o servicio (p. ej., certificado de cliente).

- ② Características:



Propiedades del Programa



① Interacción con almacenamiento en nube (p. ej., Blob Storage).

② Características:



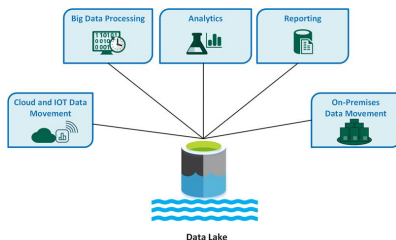
Propiedades del Programa



- ❶ Interacción con almacenamiento en nube (p. ej., Blob Storage).
- ❷ Características:

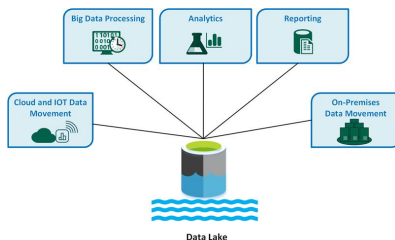


Azure Data Lake



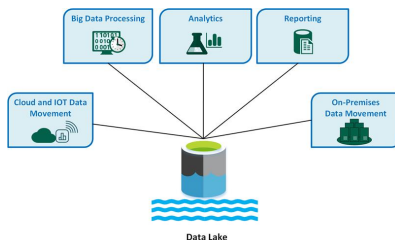
- ❶ Almacenamiento (Data Lake Storage) y análisis de datos (Data Lake Analytics).
 - *Estructurados y **No Estructurados***
- ❷ Tratamiento de grandes volúmenes en paralelo.
- ❸ Lenguajes compatibles:
 - U-SQL (SQL + C#), Python, R y Java.

Azure Data Lake



- ❶ Almacenamiento (Data Lake Storage) y análisis de datos (Data Lake Analytics).
 - *Estructurados y **No Estructurados***
- ❷ Tratamiento de grandes volúmenes en paralelo.
- ❸ Lenguajes compatibles:
 - U-SQL (SQL + C#), Python, R y Java.

Azure Data Lake

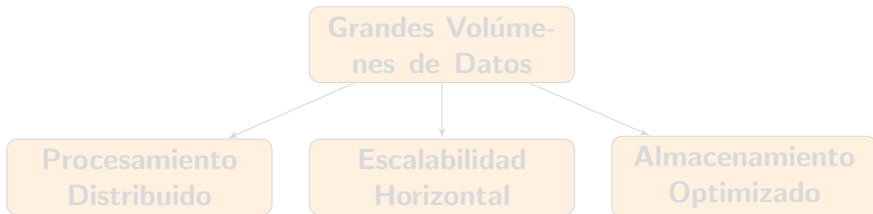


- ❶ Almacenamiento (Data Lake Storage) y análisis de datos (Data Lake Analytics).
 - *Estructurados y **No Estructurados***
- ❷ Tratamiento de grandes volúmenes en paralelo.
- ❸ Lenguajes compatibles:
 - U-SQL (SQL + C#), Python, R y Java.

Usos y Grandes Volúmenes de Datos

① Usos de Data Lake:

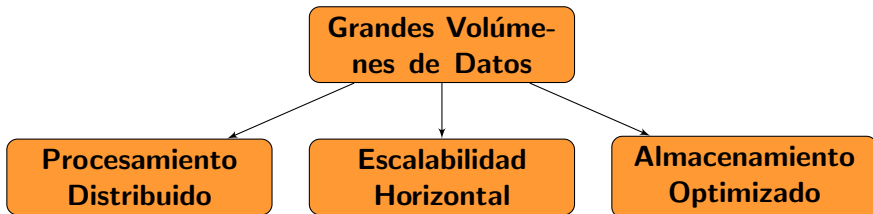
- Almacenamiento y análisis de grandes volúmenes de datos.
- Sistemas de análisis en tiempo real.
- Creación de aplicaciones para ML.



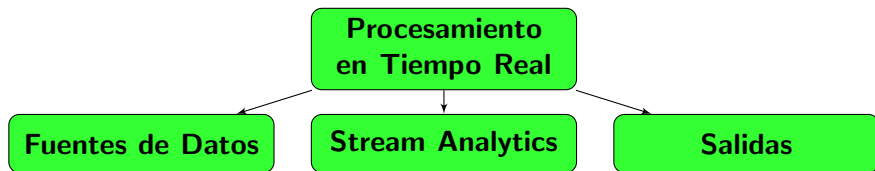
Usos y Grandes Volúmenes de Datos

① Usos de Data Lake:

- Almacenamiento y análisis de grandes volúmenes de datos.
- Sistemas de análisis en tiempo real.
- Creación de aplicaciones para ML.



Procesamiento en Tiempo Real



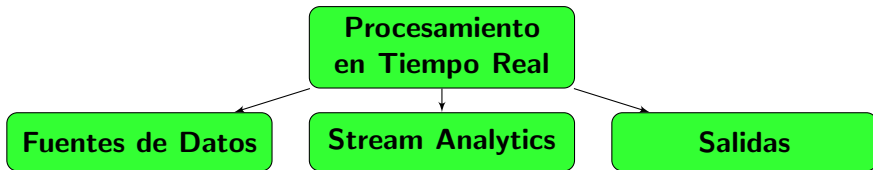
- Fuentes: Blob Storage, Data Lake Storage, Event Hubs, etc.
- Salidas: Base SQL, Blob Storage, etc.



Blob Storage:

- Almacenamiento de datos tipo documentos, imágenes, video...
- Altamente escalable.
- Integración con otros servicios de Azure.

Procesamiento en Tiempo Real



- Fuentes: Blob Storage, Data Lake Storage, Event Hubs, etc.
- Salidas: Base SQL, Blob Storage, etc.

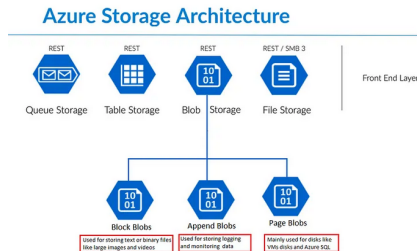
Microsoft Azure
Blob Storage



Blob Storage:

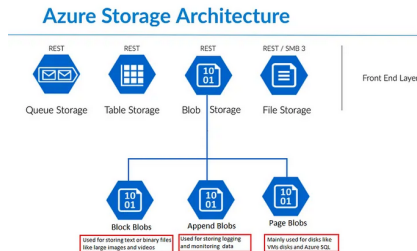
- Almacenamiento de datos tipo documentos, imágenes, video...
- Altamente escalable.
- Integración con otros servicios de Azure.

Procesamiento en Tiempo Real: Blob Storage



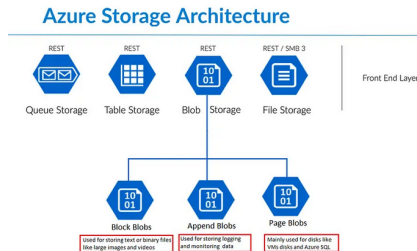
- ❶ Acceso mediante URL única.
- ❷ Altamente escalable.
- ❸ Integrable con Azure Data Factory.
- ❹ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento en Tiempo Real: Blob Storage



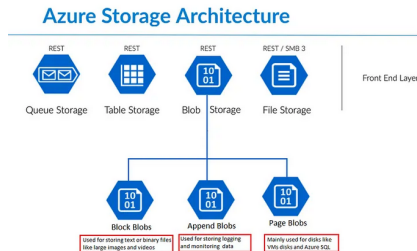
- ❶ Acceso mediante URL única.
- ❷ Altamente escalable.
- ❸ Integrable con Azure Data Factory.
- ❹ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento en Tiempo Real: Blob Storage



- ❶ Acceso mediante URL única.
- ❷ Altamente escalable.
- ❸ Integrable con Azure Data Factory.
- ❹ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento en Tiempo Real: Blob Storage



- ❶ Acceso mediante URL única.
- ❷ Altamente escalable.
- ❸ Integrable con Azure Data Factory.
- ❹ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento de Datos: Azure SQL Database



- ❶ Creación y administración de bases SQL en la nube.
- ❷ Motor de SQL Server.
- ❸ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ❹ Escalabilidad y Seguridad.

Procesamiento de Datos: Azure SQL Database



- ❶ Creación y administración de bases SQL en la nube.
- ❷ Motor de SQL Server.
- ❸ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ❹ Escalabilidad y Seguridad.

Procesamiento de Datos: Azure SQL Database



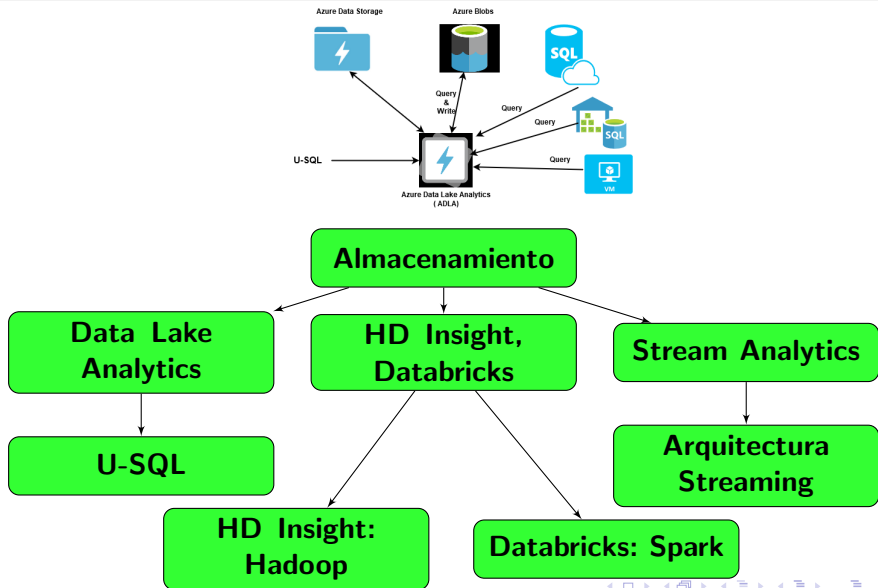
- ❶ Creación y administración de bases SQL en la nube.
- ❷ Motor de SQL Server.
- ❸ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ❹ Escalabilidad y Seguridad.

Procesamiento de Datos: Azure SQL Database

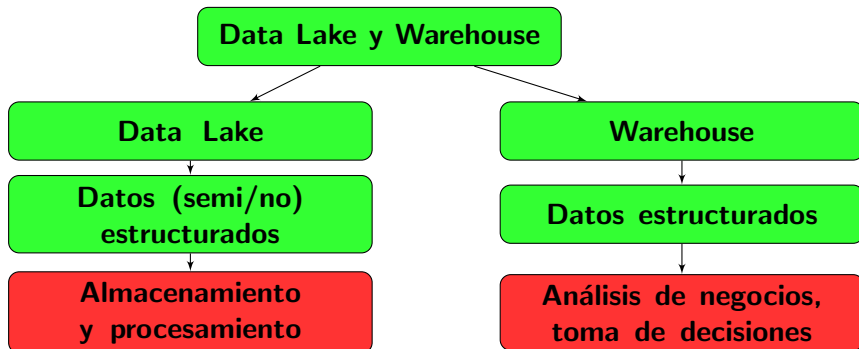
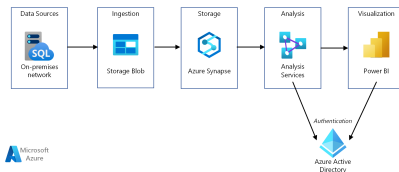


- ❶ Creación y administración de bases SQL en la nube.
- ❷ Motor de SQL Server.
- ❸ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ❹ Escalabilidad y Seguridad.

Almacenamiento en Paralelo



Data Lake y Data Warehouse



Características Principales



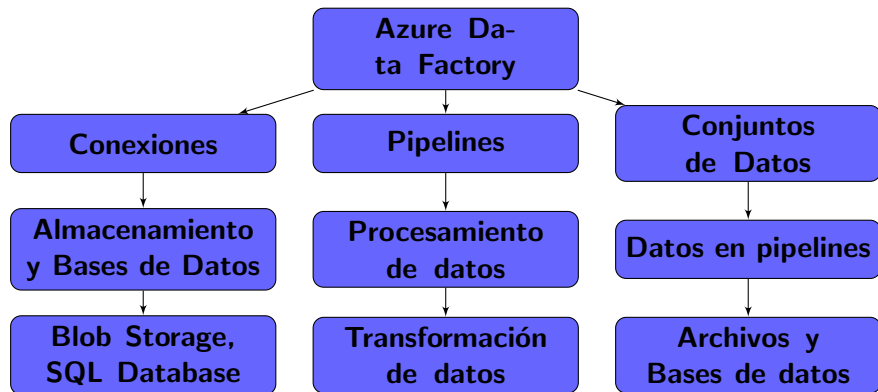
- ❶ Procesamiento de datos en nube.
 - Creación, programación y orquestación de flujos de datos.
 - Movimiento, transformación y carga.
- ❷ Integración de Azure Data Factory:
 - Servicios de Azure (Databricks, Machine Learning, Stream Analytics).
 - Visualización de datos (Power BI, Excel).

Características Principales

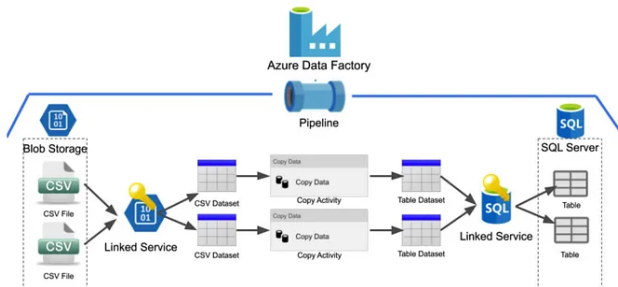


- ❶ Procesamiento de datos en nube.
 - Creación, programación y orquestación de flujos de datos.
 - Movimiento, transformación y carga.
- ❷ Integración de Azure Data Factory:
 - Servicios de Azure (Databricks, Machine Learning, Stream Analytics).
 - Visualización de datos (Power BI, Excel).

Usos

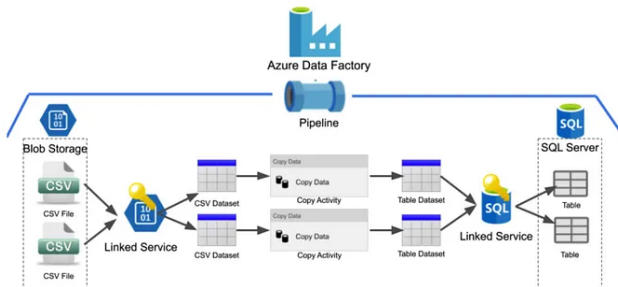


Data Factory como ETL



- ❶ Extracción: datos locales/en nube.
- ❷ Transformación/Limpieza: SQL, Python, herramientas visuales.
- ❸ Carga: Bases, Datalakes.
 - *Linked Service*: conexión a un origen/destino de datos.

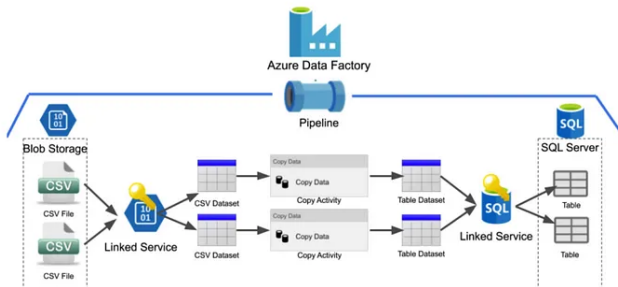
Data Factory como ETL



- ❶ Extracción: datos locales/en nube.
- ❷ Transformación/Limpieza: SQL, Python, herramientas visuales.
- ❸ Carga: Bases, Datalakes.

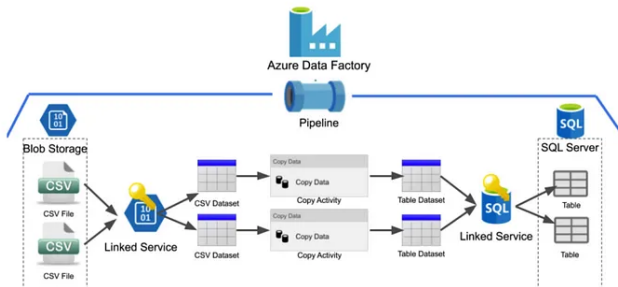
• *Linked Service*: conexión a un origen/destino de datos.

Data Factory como ETL



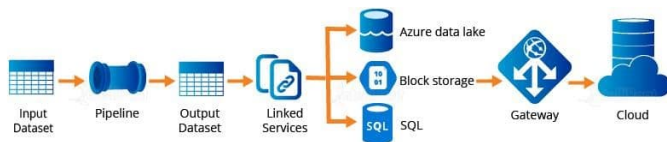
- ❶ Extracción: datos locales/en nube.
- ❷ Transformación/Limpieza: SQL, Python, herramientas visuales.
- ❸ Carga: Bases, Datalakes.
 - *Linked Service*: conexión a un origen/destino de datos.

Data Factory como ETL



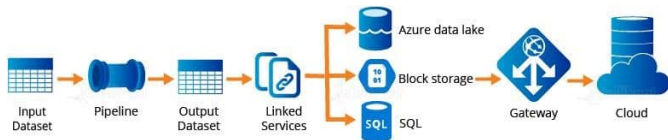
- ❶ Extracción: datos locales/en nube.
- ❷ Transformación/Limpieza: SQL, Python, herramientas visuales.
- ❸ Carga: Bases, Datalakes.
 - *Linked Service*: conexión a un origen/destino de datos.

Datasets/Conjuntos de Datos



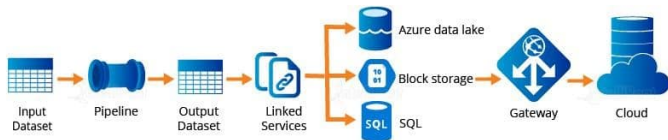
- ❶ Acceso a datos almacenados.
 - Uso en pipelines o dataflows.
- ❷ Fuentes/destinos: Blob Storage, Data Lake, SQL Server, etc.
- ❸ Formatos:
 - Archivos de texto, excel, semiestructurados (JSON, XML), SQL (SQL server, MySQL).
 - Blob storage, Data Lake storage, Azure SQL Database.

Datasets/Conjuntos de Datos



- ❶ Acceso a datos almacenados.
 - Uso en pipelines o dataflows.
- ❷ Fuentes/destinos: Blob Storage, Data Lake, SQL Server, etc.
- ❸ Formatos:
 - Archivos de texto, excel, semiestructurados (JSON, XML), SQL (SQL server, MySQL).
 - Blob storage, Data Lake storage, Azure SQL Database.

Datasets/Conjuntos de Datos



- ❶ Acceso a datos almacenados.
 - Uso en pipelines o dataflows.
- ❷ Fuentes/destinos: Blob Storage, Data Lake, SQL Server, etc.
- ❸ Formatos:
 - Archivos de texto, excel, semiestructurados (JSON, XML), SQL (SQL server, MySQL).
 - Blob storage, Data Lake storage, Azure SQL Database.

Pipelines y Dataflows



Pipelines:

- Orquestación de procesos ETL.
- Operaciones: dataflows, mover y transformar, anexar variables, borrar, etc.



Dataflows:

- Procesamiento y transformación de grandes volúmenes de datos.
- Transformación datos entrada (interfaz gráfica).
- Alimento de Python y SQL.

Pipelines y Dataflows



❏ Pipelines:

- Orquestación de procesos ETL.
- Operaciones: dataflows, mover y transformar, anexar variables, borrar, etc.

❏ *Dataflows:*

- Procesamiento y transformación de grandes volúmenes de datos.
- Transformación datos entrada (interfaz gráfica).
 - A través de Python y SQL.

Actividades en Pipelines

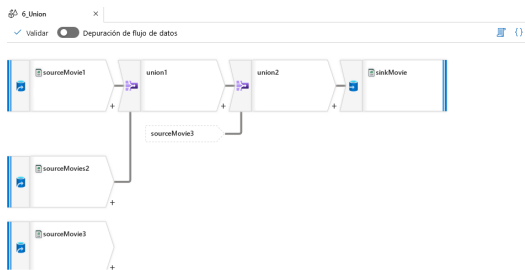
- ❶ *Mover y Transformar:*
 - Copiar datos, Dataflow (basado en SQL y Spark).
- ❷ *Actividades Generales:*
 - If, ForEach, Set Variable, Until, Switch.
- ❸ *Actividades de Data Flow:*
 - Source, Derived Column, Filter, Aggregates, Join, Sink.

Actividades en Pipelines

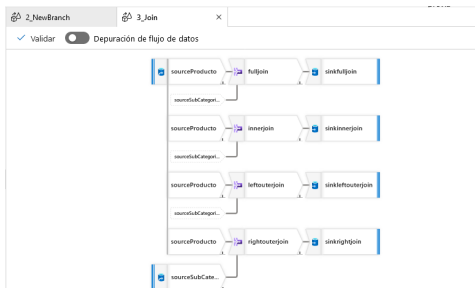
- ❶ *Mover y Transformar:*
 - Copiar datos, Dataflow (basado en SQL y Spark).
- ❷ *Actividades Generales:*
 - If, ForEach, Set Variable, Until, Switch.
- ❸ *Actividades de Data Flow:*
 - Source, Derived Column, Filter, Aggregates, Join, Sink.

Actividades en Pipelines

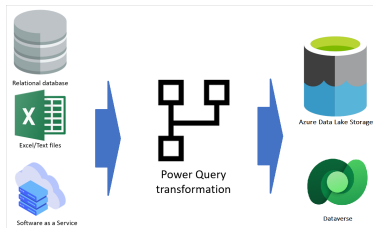
- 🔊 *Mover y Transformar:*
 - Copiar datos, Dataflow (basado en SQL y Spark).
- ⏸ *Actividades Generales:*
 - If, ForEach, Set Variable, Until, Switch.
- 🔊 *Actividades de Data Flow:*
 - Source, Derived Column, Filter, Aggregates, Join, Sink.



Dataflow en Azure Data Factory



Power Query



❶ Actividad de transformación de datos.

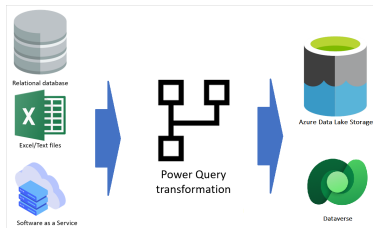
- Dentro de un Pipeline.
- Fuentes: archivos, DBs, APIs.

❷ Uso:

- Definir actividad Power Query en Pipeline.

• Definir Pipeline para la transformación y el movimiento de datos.

Power Query



- ❶ Actividad de transformación de datos.
 - Dentro de un Pipeline.
 - Fuentes: archivos, DBs, APIs.
- ❷ Uso:
 - Definir actividad Power Query en Pipeline.

- Ejemplos: Limpieza, transformación y agrupación de datos.